J. Mol. Biol. (1984) 180, 549–576

Annex 5

Translation is a Non-uniform Process

Effect of tRNA Availability on the Rate of Elongation of Nascent Polypeptide Chains

Stanislas Varenne, Jean Buc, Roland Lloubes and Claude Lazdunski

Centre de Biochimie et de Biologie Moléculaire du C.N.R.S.
B.P. 71, 13402 Marseille Cedex 9, France

(Received 2 December 1983, and in revised form 18 July 1984)

We reported elsewhere (Varenne *et al.*, 1982) that, during synthesis of a number of colicins in *Escherichia coli*, intermediate nascent chains of discrete sizes accumulated, suggesting a variable rate of translation. In this paper, a detailed analysis provides arguments that this phenomenon, at least for the proteins under study, is not related to aspects of messenger RNA such as secondary structure. It is linked to the difference in transfer RNA availability for the various codons. Experimental analysis of translation of other proteins in *E. coli* confirms that the main origin for the discontinuous translation in the polypeptide elongation cycle is the following. For a given codon, the stochastic search of the cognate ternary complex (aminoacyl-tRNA–EF-Tu-GTP) is the rate-limiting step in the elongation cycle: transpeptidation and translocation steps are much faster. The degree of slackening in ribosome movement is almost proportional to the inverse of tRNA concentrations. The verification of this model and its possible physiological significance are discussed.

1. Introduction

The idea that intracellular concentrations of tRNAs play an important role in the dynamics and the regulation of protein synthesis was suggested 20 years ago by Ames & Hartman (1963). Anderson (1969) proposed that the rate of translation might be slowed *in vivo* at the site of regulatory codons. Since then, more detailed knowledge of nucleotide sequences, tRNA concentrations, decoding spectra of tRNA species, energies of interaction between codons and anticodons, has led to a number of authors emphasizing the importance of tRNA concentrations and/or the role of codon–anticodon interaction energies in the dynamics of translation (e.g. see Grantham *et al.*, 1981; Chavancy & Garel, 1981; Gouy & Gautier, 1982; Ikemura, 1981*a*,*b*,1982; Grosjean & Fiers, 1982). The idea that translation occurs at a variable rate is implicit or explicit in their work, but to our knowledge, no study has yet proposed a quantitative relationship between tRNA concentration and rate of elongation at each codon.

It has been observed in our laboratory that elongation of nascent polypeptide chains of colicin A, E1, E2 and E3 occurs at a variable rate (Varenne *et al.*, 1982;

0022-2836/84/330549-28 $03.00/0

Lazdunski *et al.*, 1984). Since the nucleotide sequence for colicin A has been established (Morlon *et al.*, 1983), and since colicin A is highly expressed after induction, whereas synthesis of chromosomal proteins is strongly reduced, colicin A synthesis constitutes a good model system for an experimental approach to the dynamics of the elongation cycle.

In this work, we have analysed in detail the effect of tRNA availability on the polypeptide elongation rate for colicin A and for some other *Escherichia coli* proteins.

Our results demonstrate directly two important points: (1) the rate-limiting step in the elongation cycle of polypeptide is tRNA selection; (2) as the different tRNAs exist in *E. coli* in different intracellular concentrations (even those preferentially used by highly expressed proteins), the rate of translation varies along the mRNA. Moreover, for a given protein, complete translation of individual mRNAs occurs at different rates.

The possible effect of codon–anticodon energies of interaction has not been taken into account for lack of adequate data. However, our results indicate that the role of this factor in the rate of translation, if it existed, would have to be less important than that played by tRNA concentrations.

2. Materials and Methods

(a) *Materials, bacterial strains, growth conditions and conditions of radiolabelling*

All materials were as described (Varenne *et al.*, 1982). Procedures were also carried out exactly as described (Varenne *et al.*, 1982). For production of colicin E1, the strain *E. coli* K12 W3110 ColE1, *nal*[r] was used.

(b) *Preparation of samples for immunoprecipitation*

In addition to the previously reported experimental protocol, another technique was occasionally used to solubilize cell pellets. These were taken up in 10 μl of 2% (w/v) sodium dodecyl sulphate and incubated for 5 min at 100°C, allowed to cool slowly and 90 μl of the immunoprecipitation buffer (without sodium dodecyl sulphate) added as described (Varenne *et al.*, 1982).

(c) *Antibodies, gel electrophoresis and fluorography*

Antibodies and antisera were obtained as described (Varenne *et al.*, 1982). Gel electrophoresis and fluorography were also carried out as described (Varenne *et al.*, 1982), except for a pre-exposure of the film to a hypersensitizing light flash resulting in an increase of background film absorbance of about 0·15 at 540 nm. This treatment permits quantitative interpretation of film density (Laskey & Mills, 1975), using, if necessary, different exposures for very contrasted fluorograms.

(d) *Analysis of polypeptide elongation cycle: the stochastic model*

Gouy & Grantham (1980), Chavancy & Garel (1981) and Gouy (1981) have pointed out the influence of tRNA availability on the dynamics of protein synthesis. The elongation cycle is described by Gouy & Gautier (1982) thus:

"At the beginning of each polypeptide elongation cycle, a codon is found at the ribosomal A-site. The ternary complexes (aminoacyl-tRNAs bound with elongation factor Tu and GTP) diffusing in the cytoplasm interact with the codon and the ribosome at the A-site. Most often the codon does not belong to the tRNA recognition spectrum and therefore the aminoacyl-tRNA dissociates from the ribosome. When the specificity condition is fulfilled, the elongation cycle starts: transpeptidation and translocation occur. Thus, each codon can be characterized by the average number of codon–tRNA interactions at the A-site during one elongation cycle. The relative concentration of the codon–cognate tRNA is equivalent to its probability of colliding with the A-site codon. Hence, if this probability is C the mean number of codon–tRNA interactions necessary for the elongation cycle to occur is $1/C$."

The relationship between t, the mean duration for the addition of a given amino acid residue corresponding to a given codon, and $1/C$ ($1/C$ will be called N) is the following: t is the sum of 3 mean durations: t_1 for the search of the adapted ternary complex for the codon in the A-site of the ribosome; t_2 for transpeptidation; and t_3 for translocation. Let θ_0 be the mean duration of an interaction between a given codon and a non-cognate tRNA at the A-site of the ribosome, and let θ_1 be the mean duration between the ejection of a non-cognate tRNA from the A-site and the next collision with a tRNA. Then the mean duration of the addition of a given amino acid residue corresponding to a given codon is $t = (\theta_0 + \theta_1)N + t_2 + t_3$. θ_1 is independent of the codon considered. If we assume that θ_0, t_2 and t_3 are also independent of the codon, then $t = AN + B$, where A and B are constants.

(e) *Determination of tRNA concentrations used for computation of N*

Since concentrations of only some tRNA species have been determined experimentally, it was necessary to calculate an approximate value for the remainder. We have taken advantage of the fact that a good correlation exists between the abundance of tRNAs and their usage in the cell (Ikemura, 1981*a*,*b*; Gouy, 1981). It is thus possible to determine by interpolation an approximate value of unknown tRNA concentrations by using the linear regression plot of experimentally determined tRNA concentrations as a function of usage of these tRNAs in total *E. coli* proteins. This general usage can be approximated from the particular usage for a number of mRNAs for which sequences are known. The degree of correlation between the abundance of tRNA species and the usage of the supposedly corresponding codons is more or less strong according to whether the particular protein is highly or weakly expressed. Bearing this in mind, Gouy (1981) has established an average tRNA usage by considering that 60% of total proteins can be considered as highly expressed and 40% as weakly expressed.

In each class, each protein is involved only by the number of times that each of the 61 codons appears in its sequence and not by its amount in the cell. However, this is not of overriding importance since, in each class, each protein uses tRNAs in approximately the same way (Ikemura, 1981*a*,*b*). On the same grounds, we have established a new codon usage (Table 1), which differs slightly from that established by Gouy (1981) for 2 reasons: (1) extrachromosomal genes were not taken into account, since these genes are not expressed permanently in *E. coli*; (2) numerous new sequences have been determined and we have used 21 sequences, or part sequences, corresponding to highly expressed genes (Gouy & Gautier, 1982), and 41 sequences of weakly expressed genes. The codon usage for the latter comes from the Lyon sequence bank ACNUC. Among these 41 genes, 28 are referred to by Gouy & Gautier (1982). They are: *ompC*, *lacI*, *lacY*, *ilvG*, *rpoA*, *rpoB*, *rpoD*, *trpA*, *trpB*, *trpC*, *trpD*, *trpE*, *trpR*, *thrA*, *uncI*, *uncB*, *uncE*, *uncF*, *uncH*, *uncA*, *uncC*, *uncD*, *uncG*, *araC*, *aroH*, *lexA*, *ndh*, *asnA*. The 13 new genes are the following: *alaS* (Putney *et al.*, 1981), *crp* (Cossart & Gicquel-Sanzey, 1982; Aiba *et al.*, 1982), *fol* (Smith & Calvo, 1980), *galR* (Von Wilcken-Bergmann & Muller-Hill, 1982), *lamB* (Clément & Hofnung, 1981), *polA* (Joyce *et al.*, 1982), *rpoC* (Ovchinnikov *et al.*, 1982), *thrB* and *thrC* (Cossart *et al.*, 1981), *tnaA* (Deeley & Yanofsky, 1981), *trpS* (Hall *et al.*, 1982), 13 Kd and 15 Kd proteins (Grundström & Jaurin, 1982).

The codon usage corresponding to the 62 genes has been allocated between tRNAs by using the decoding spectrum proposed by Ikemura (1981a). For codons recognized by 2 iso-tRNAs, the repartition between these 2 tRNAs has been carried out as indicated by Ikemura (1981a). The correlation between the amounts of tRNAs and their frequency of usage for 23 tRNAs quantified by Ikemura (1981a) is shown in Fig. 3. The correlation coefficient of 0·86 indicates that one can roughly estimate the other tRNA concentrations by interpolation from the regression line corresponding to the plot of amount of tRNA *versus* frequency of tRNA usage. The values of tRNA concentrations obtained and used in further calculations are indicated in Table 2. Thirty-five concentrations are actual concentrations (experimentally determined or interpolated). Including the $tRNA_f^{Met}$ concentrations, the total is equal to 1. The 6 concentrations concerning codons UCU (Ser), GGC (Gly), ACU (Thr), CCU (Pro), GCU (Ala) and GUU (Val) are apparent concentrations that take into account the recognition of these codons by 2 different tRNAs.

In a number of calculations, the value for the $tRNA^{Ile}_{minor}$ has been replaced by the interpolated value. The latter was obtained from the regression line derived from only 22 points, but was very close to that obtained with 23 points.

Grosjean & Fiers (1982) proposed a decoding spectrum that differs from that of Ikemura (1981a) for 6 codons: (1) according to these authors, 2 different tRNA Leu iso-acceptors translate codons UUA and UUG: as the codon usages are quite similar (about 50% *versus* 50%), the experimentally determined concentration has been distributed equally between these 2 tRNAs; (2) according to Grosjean & Fiers (1982), it is not certain that codons AAG (Lys) and GAG (Glu) can be decoded by the tRNAs decoding AAA (Lys) and GAA (Glu), respectively. In this hypothesis, we have assumed that experimentally determined concentrations concerned the sum of the 2 tRNAs, and these concentrations have been distributed between these 2 tRNAs proportionally to the codon usage. In the hypothesis where the experimentally determined value was related to the major tRNA, the concentrations should be modified accordingly, but this has minor consequences for further calculation.

(f) Data treatment: determination of the average number of codon–tRNA interactions at the ribosome A-site during one elongation cycle (abbreviated to N) as a function of the migration of the corresponding elongation intermediates

Basic programs were developed with a Wang 2200 microcomputer provided with a floppy disk and a digital plotter.

As a first step, fluorograms obtained (10 cm migration) were enlarged 5 times under conditions that preserve contrast. The enlargement was scanned with a 4 times expansion in migration. Knowing the migration of calibration standards, one can determine ~~coefficients a and b for~~ the migration according to $x = a \log M_r + b$.

For each codon (C_i) the computer determines the cumulated weight of all amino acid residues assembled until the corresponding amino acid (A_i), the corresponding position in the scanning (x_i) and the number N_i ($1/C_i$, where C_i is the frequency of the tRNA specific for codon i, routinely called concentration for convenience). After computation, the plot x_i (abscissa), N_i (ordinate) is drawn. As a certain dispersion in densitometer profiles exists (diffusion of polypeptides in the gel, light-scattering in the fluorogram and slit width), it is necessary to simulate this diffusion. This dispersion can be accomplished through a Gaussian distribution with a full width at half maximum (FWHM) either constant in distance ($\Delta x = cte$), or constant in M_r ($\Delta M_r = cte$). A convolution (called dispersion in the text and denoted N^*) from the direct calculations is thus obtained.

Another computer program plots x_i (abscissa), N_i (ordinate) with $x_i = C(i+d)$, and disperses the results with a full width at half maximum constant in amino acid residue number ($\Delta i = cte$). For a search of regions of the mRNA where collisions between 2 adjacent ribosomes might occur, the Gaussian distribution is replaced by a unit distribution with a width of n amino acid residues.

TABLE 1

Average codon usage in E. coli

Arg CGA	1		Ser AGC	10		Gly GGG	4		Asp GAC	31
CGC	19		AGU	4		GGU	41		GAU	21
CGG	1		Thr ACA	4		Val GUA	22		Tyr UAC	15
CGU	38		ACC	25		GUC	9		UAU	7
AGA	1		ACG	5		GUG	18		Cys UGC	4
AGG	0		ACU	20		GUU	32		UGU	3
Leu CUA	1		Pro CCA	7		Lys AAA	52		Phe UUC	20
CUC	5		CCC	2		AAG	20		UUU	10
CUG	62		CCG	25		Asn AAC	31		Ile AUA	0
CUU	5		CCU	3		AAU	6		AUC	41
UUA	5		Ala GCA	28		Gln CAA	9		AUU	17
UUG	6		GCC	14		CAG	31		Met AUG	23
Ser UCA	3		GCG	28		His CAC	13		Trp UGG	7
UCC	14		GCU	46		CAU	7			
UCG	3		Gly GGA	3		Glu GAA	50			
UCU	17		GGC	31		GAG	16			

Determination of this codon usage has been carried out as described in Materials and Methods. Values are expressed per thousand and are approximated to the next integer.

TABLE 2

Values of tRNA concentrations used in the calculation of the average number of selections (N) expressed as %

Amino acid	Codon	I[a]	GF[a]	Amino acid	Codon	I[a]	GF[a]
Arg	CG(U, C, A)	5·05		Val	GUU[e]	4·01[c]	
	CGG	0·68[b]			GUC	2·24	
	AG(A, G)	0·68[b]			GU(A, G)	5·80	
Leu	CU(U, C)	1·68		Lys	AAA	5·61	4·04[d]
	CUA	0·68[b]			AAG		1·57[d]
	CUG	5·61		Asn	AA(U, C)	3·37	
	UUA	1·40	0·70[d]	Gln	CAA	1·68	
	UUG		0·70[d]		CAG	3·24	
Ser	UCU[e]	3·20[bc]		His	CA(U, C)	2·21	
	UCC	2·78[b]		Glu	GAA	5·05	3·74[d]
	UC(A, G)	1·40			GAG		1·31[d]
	AG(U, C)	1·40		Asp	GA(U, C)	4·49	
Thr	ACU[e]	5·01[c]		Tyr	UA(U, C)	2·90	
	ACC	4·49		Cys	UG(U, C)	1·43[b]	
	AC(A, G)	1·73[b]		Phe	UU(U, C)	1·08	
Pro	CCU[e]	1·82[bc]		Ile	AU(U, C)	5·61	
	CCC	0·00[bc]			AUA[e]	0·29	
	CC(A, G)	3·08[c]			AUA[e]	0·60[f]	
Ala	GCU[e]	3·80[bc]		Met	AUG	1·03	
	GCC	4·07[bc]		Trp	UGG	1·68	
	GC(A, G)	5·77[b]					
Gly	GG(U, C)	6·17					
	GGA	0·84					
	GGC[e]	1·40[c]					

[a] Recognition pattern according to (I) Ikemura (1981a,b); (GF) Grosjean & Fiers (1982).
[b] Interpolated values.
[c] Recognition of 1 codon by 2 tRNAs was taken into account.
[d] Distributed values.
[e] Apparent concentration.
[f] Two different possible values for concentration (see Materials and Methods).

3. Results

Since our previous report of a non-uniform rate of translation for some colicin mRNAs (Varenne *et al.*, 1982), the nucleotide sequences of colicin El (Yamada *et al.*, 1982) and colicin A (Morlon *et al.*, 1983) have been established. Since frequent usage of codons corresponding to minor iso-tRNAs of *E. coli* was observed (Morlon *et al.*, 1983), it became more likely that codon usage, rather than mRNA secondary structure, was the predominant factor in the mechanism responsible for creating pauses in elongation.

A detailed analysis of discrete intermediates in the elongation of nascent polypeptide chains required the establishment of experimental conditions leading to results reflecting as accurately as possible the true concentrations of the intermediates.

(a) *Detailed analysis of colicin A intermediates*

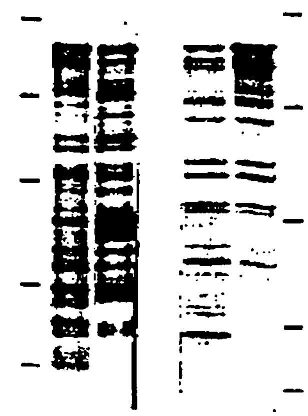
The main problems concerning quantitation of the nascent polypeptide chains were as follows: (1) their radiolabelling should be as uniform as possible for all M_r values; (2) proteolysis should be avoided or at least minimized; (3) the recovery of polypeptides in the solubilization process should be constant for all intermediate sizes; and (4) the yield of immunoprecipitation should also be constant.

In analysis of colicin A intermediates, the following conditions were used.

(1) Pulses of 20 to 25 seconds with [^{35}S]methionine allowed a rather uniform labelling in spite of the unequal distribution of methionine residues along the polypeptide chain. This unexpected result, which is analysed in the Discussion, was deduced from the comparison of fluorograms (not shown) obtained from total cell proteins of the fully induced strain CA31 ColA radiolabelled with a ^{14}C-labelled amino acid mixture or with [^{35}S]methionine. Under these conditions, colicin A represents more than 50% of total protein (Varenne *et al.*, 1981) and a direct comparison of many intermediates is possible.

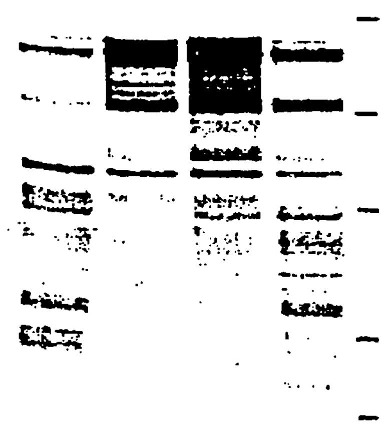
(2) It is well known that incomplete polypeptide chains made by nonsense mutant strains of *E. coli* are degraded at different rates, which are not directly proportional to their length but are apparently determined by their conformation (Lin & Zabin, 1972). The same phenomenon could be observed for nascent chains of colicin A *in vitro* (Varenne *et al.*, 1981) or *in vivo* (Fig. 1(a), lane 2). Comparison of fluorograms of trichloroacetic acid precipitates and immunoprecipitates of fully induced cells (Varenne *et al.*, 1982) showed that with the experimental conditions described under Materials and Methods, this proteolysis was generally weak for colicin A intermediates.

(3) and (4) The best conditions for solubilization and immunoprecipitation were determined. When a very efficient solubilization treatment of cell membranes was performed (Fig. 2, lane 3), all the intermediates and the terminated colicin A were present in the immunoprecipitation mixture, but considerable inhibition of low M_r intermediate immunoprecipitation was apparent, due to a limiting antibody-antigen ratio (an increase of this ratio led to perturbations in the pattern of intermediates and must be avoided). The low M_r intermediates that escaped



1 2 3 4

Fig. 1. Effect of proteolysis and solubilization technique in recovery of colicin A intermediates. Fully induced cells were pulse-labelled for 20 s with [^{35}S]methionine and chloramphenicol (200 μg/ml) was added. One sample (lane 1) was immediately solubilized and immunoprecipitated; another (lane 2) was first incubated at 37°C for 1 h, then submitted to the same treatment. One sample (lane 3) was solubilized as described (Varenne *et al.*, 1982) and a second (lane 4) was solubilized as indicated in Materials and Methods. Molecular weight standards: bovine serum albumin, 67,000 M_r; ovalbumin, 46,000 M_r; carbonic anhydrase, 30,000 M_r; soybean trypsin inhibitor, 20,100 M_r; lysozyme, 14,400 M_r.



1 2 3 4

Fig. 2. Consequence of binding of large intermediate nascent polypeptide chains of colicin A to cell membrane in the solubilization process. Fully induced cells were pulse-labelled for 20 s and solubilized before immunoprecipitation either (lane 4) as described (Varenne *et al.*, 1982) or (lane 3) as indicated in Materials and Methods. The supernatant of the latter immunoprecipitate was submitted to an additional immunoprecipitation (lane 1). Pellet recovered after the mild solubilization was solubilized through the harsh process and immunoprecipitation was again carried out (lane 2).

immunoprecipitation could be recovered in the supernatant and were totally immunoprecipitated in a second step procedure (Fig. 2, lane 1). Conditions that cause an incomplete solubilization of the membrane fraction (Fig. 2, lane 4) allowed an efficient immunoprecipitation of low M_r intermediates because much less competition by mature colicin A then occurred. Missing intermediates of high M_r and colicin A could be recovered (lane 2) from the membrane pellet with a harsher solubilization process similar to that used previously (lane 3).

By combining densitometer profiles from fluorograms obtained under both conditions in another experiment (Fig. 1(b), lanes 3 and 4), corrected for low M_r intermediates by quantification of these intermediates in whole fully induced cells (not shown), a profile was obtained that reflected the real intensities of intermediates (Fig. 5(a)).

Thus, experimentally we have access to t, the mean duration of addition of a given acid residue corresponding to a given codon. If the tRNA concentration in colicin A-producing cells were known, it would then be possible to check if the non-variable rate of elongation might be explained by the stochastic model described in Materials and Methods.

(b) *The internal pool of tRNAs is not perturbed upon synthesis of colicin A*

Intracellular tRNA concentrations for *E. coli* have been determined by various authors and particularly by Ikemura (1981*a*). In order to use these tRNA concentrations in our calculations, we had to make sure that the internal pool of various tRNAs was not perturbed upon synthesis of a very highly expressed protein like colicin A that displays an unusual tRNA usage as shown in Figure 3. To allow an easy comparison with other proteins examined by Ikemura (1981*a*,*b*) in *E. coli*, the data points were analysed by linear regression. The regression line is expressed by $y = ax+b$. The amount of tRNA $= x$, the frequency of tRNA usage $= y$, the correlation coefficient $= r$:

with 19 tRNA concentrations, $y = 4{\cdot}2x+0{\cdot}80$, and $r = 0{\cdot}65$,
with 23 tRNA concentrations, $y = 3{\cdot}7x+1{\cdot}12$, and $r = 0{\cdot}63$.

These results compared with those of Ikemura (1981*a*) clearly show that tRNA usage of colicin A is different from highly or weakly expressed proteins, and more generally different from the mean usage of *E. coli* proteins (see Fig. 3), for which $y = 6{\cdot}9x-0{\cdot}48$ and $r = 0{\cdot}96$.

As previously emphasized (Varenne *et al.*, 1982), an increase of about 100-fold in colicin A synthesis by induction had no effect on the intermediate intensities, but mitomycin C only increased the number of induced cells and did not modify the amount of colicin A produced by each induced cell. Therefore, this stability of intermediates after induction did not allow any conclusions to be drawn as to the effect of the amount of colicin A produced in each cell. In order to clarify this point, cells were pulsed and chased at different times after mitomycin C addition (Fig. 4, lanes 1 to 9).

Under the conditions used, all cells were induced after 15 minutes of incubation

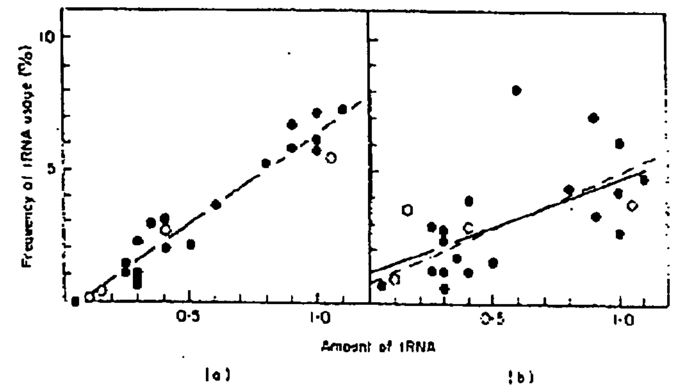


Fig. 3. The relationship between tRNA abundance and its usage found for *E. coli* genes and for *caa* (gene of colicin A). For the reasons indicated by Ikemura (1981*a*), data for $tRNA_f^{Met}$, Ala1, Thr3+1 and Ser1 are not shown. Data for Gly1, Gly2, Val1 and Val2 are specified by open circles (see the text). Linear regression analysis was performed with 23 tRNAs (continuous lines) and with only 19 tRNAs for colicin A (broken line) to facilitate comparison with individual proteins analysed by Ikemura (1981*a*,*b*). (a) *E. coli*; (b) colicin A.

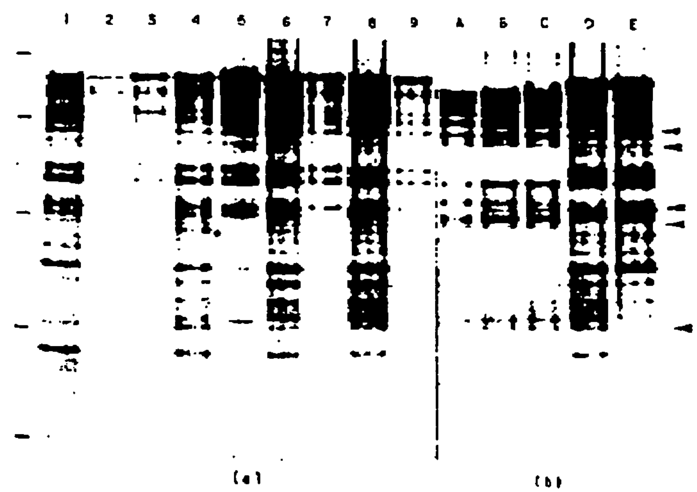


Fig. 4. Intermediates are not modified by overproduction of colicin A. Positioning of intermediates. (a) Cells were pulse-labelled either before induction (lanes 2 and 3) or after induction of 15 min (lanes 4 and 5), 55 min (lanes 6 and 7), or 100 min (lanes 8 and 9). In lane 1, the sample applied was similar to that applied in lane 8 but for the solubilization (see the text). Lanes 7 and 9 were intentionally underexposed to allow detection of the stage of chase. (b) Cells were pulse-labelled at 25°C with (^{35}S)methionine for 4 s (lane A), 6 s (lane B), 8 s (lane C). Lanes D and E correspond to the chase of the 8 s pulse for 30 s and 60 s, respectively. Arrows indicate intermediates just downstream from methionine residues 185, 202, 205, 387 and 448.

with mitomycin C, and the intermediates observed (lanes 4 and 5) were essentially produced by newly induced cells (compare to lanes 2 and 3, which show colicin production by naturally induced cells). After 55 minutes (our routine induction time: lanes 6 and 7), tenfold more colicin per cell was produced than after 15 minutes of induction, and about fivefold more was produced after 100 minutes (lanes 8 and 9). In all cases, accumulated intermediates were located at the same M_r and had the same relative concentrations. The artefactual differences observed from top to bottom between lanes 4 and 6 may be explained by the low yield of immunoprecipitation of low M_r intermediates when amounts of complete colicin A become too large, as shown in Figure 2. Two lines of evidence demonstrate this point: (1) the chase was similar in lanes 5, 7 and 9; (2) when a milder cell solubilization method was applied to cells induced for 55 minutes, profiles obtained after 15 minutes (lane 4) and 55 minutes (lane 1) were directly comparable between 13,000 M_r and 48,000 M_r, that is in the M_r range where the recovery of intermediates in the solubilization process was equivalent.

This experiment clearly indicates that accumulations of intermediates (reflecting "pauses") routinely observed when colicin A is highly expressed do not result from perturbations in tRNA concentrations induced by the overproduction of colicin A. Furthermore, the experimental results in which the *Citrobacter freundii* strain CA31 ColA was used could be compared with theoretical predictions obtained based on *E. coli* tRNA concentrations, since we observed the same pattern of intermediates when the plasmid pColA was introduced into *E. coli* K12 W3110 (Varenne *et al.*, 1982).

(c) *Correlation between theoretical and experimental profiles for colicin A*

The theoretical profile corresponding to $t = AN + B$ *versus* the position of intermediates could not be plotted, since A and B are not known; however, it is possible to plot N. Then, maxima and minima of N correspond to maxima and minima of t and, if the model was correct, must correspond to maxima and minima of the experimental profile. Indeed, the amount of nascent polypeptide chains comprising n amino acid residues at the P-site of the ribosome is proportional to the mean duration of addition of the $n+1$ amino acid residue.

Owing to dispersion (see Materials and Methods), correlation was sought between the scans of intermediates (Fig. 5(b)) and the plot of N^* (the dispersed values of N: see Materials and Methods) as a function of electrophoretic migration. The most visible peaks of the fluorogram were numbered from 1 (complete colicin A) to 31 (last visible intermediate).

We observed that it was not possible to optimize the dispersion of N (a part of the N plot is shown in Fig. 6) over the whole range of polypeptide sizes, even by using a constant dispersion in M_r (ΔM_r = constant), and two different dispersions N^* with different full width at half height with Δx = constant had to be used. The best fit between theoretical and experimental profiles was obtained with the dispersion shown in Figure 5(c) for the high M_r range, and in Figure 5(a) for the low M_r range.

Since use of the experimental value for $tRNA^{Ile}_{minor}$ concentration resulted in an

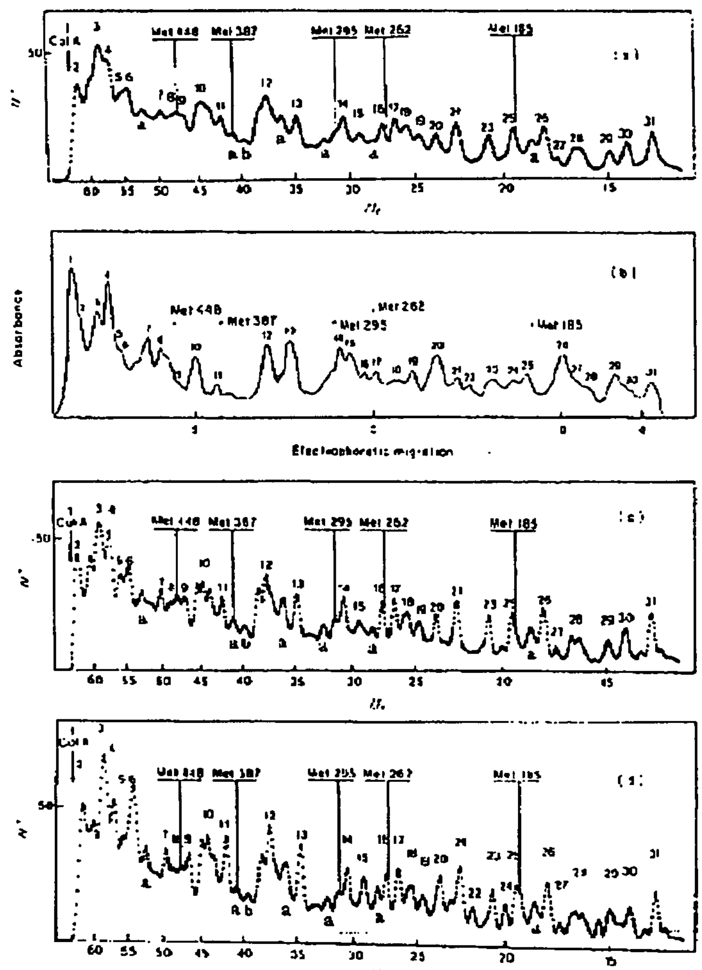


FIG. 5. Comparison of theoretical and experimental profiles for intermediates. (a) N^* (see Materials and Methods) has been plotted as a function of polypeptide chain M_r with a full width at half maximum (FWHM) = 11·6 mm. The distance between peaks 1 and 31 was 580 mm. (b) Densitometer profile for colicin A intermediates. Cells were pulse-labelled for 20 s and enlarged fluorograms from Fig. 1(b) were scanned (see the text for details). Electrophoretic migration was from right to left. Letters a, b, c and d indicate migrations of calibration proteins (a = 14,400 M_r, b = 20,000 M_r, c = 30,000 M_r and d = 40,000 M_r). Arrowheads: ▶ Met 448, ▶ Met 387, etc. indicate that methionines in positions 448, 387, etc., are located between the indicated position and the next upstream intermediate (not visible in the pulse but visible in the chase). (c) Same as (a) with FWHM = 8·4 mm. (d) Same as (c) but the decoding spectrum of tRNAs was that of Grosjean & Fiers (1982). The numbers of the amino acids (Morlon *et al.*, 1983) corresponding to the tops of the main theoretical peaks in the N^* profile are indicated in parentheses. 2 (557), 3 (549), 4 (533), 5 (518), 6 (506), 7 (467), 8 (457), 9 (438), 10 (419), 11 (397), 12 (355), 13 (326), 14 (290), 15 (279), 16 (264), 17 (263), 18 (245), 19 (237), 20 (220), 21 (215), 22 (207), 23 (196), 24 (189), 25 (186), 26 (172), 27 (166), 28 (160), 29 (146), 30 (140), 31 (131).

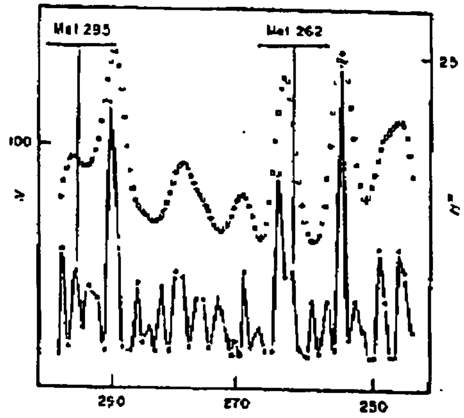


FIG. 6. Plotting of N and N^* for part of the colicin A sequence centred around amino acid residue 270. In the N profile: (✱) computed from experimental tRNA concentrations; (O) computed from interpolated tRNA concentrations. The N^* profile is a part of that shown in Fig. 5(e).

N value of 345, and for N^* profiles caused accentuated peaks at positions 255, 425, 533 and 567 (not shown) that did not have an equivalent counterpart in densitometer profiles, we suspected that the experimental value could largely be underestimated. In contrast, when the interpolated value was used (then $N = 167$), the peaks cited above were more equivalent to experimental peaks. This value was therefore used routinely for further calculations.

Although the observed theoretical profiles A and C do not exactly reproduce the experimental profile B, there is generally a good correlation between positions of theoretical and experimental peaks. It is also quite significant that regions where N^* is low (between peaks 11 and 12 and between peaks 13 and 14) contain a lower amount of intermediates than elsewhere. Some minor theoretical peaks (9a, 11a, 11b, 13a) correspond to slight bands on the fluorogram, one (25a) does not appear in this experiment but is just visible in some others, two (12a and 15a) are never observed. In contrast, the experimental peak 1 does not appear in the theoretical profiles, since it corresponds to the colicin A itself and not to an intermediate, and the experimental peaks 22 and 24 are not predicted with Ikemura's (1981a) recognition pattern. As this pattern was slightly different from that reported by Grosjean & Fiers (1982), see Materials and Methods, it was important to check the effect of this difference in the above correlation. In the new pattern obtained (Fig. 5(d)), the position for most of the peaks remains unmodified although areas are locally changed. The main difference can be observed for peaks 22 and 24 that now appear.

This feature is particularly interesting, because the modifications in these two peaks result only from the existence of three GAA-GAG-GAA codons for glutamic acid in positions 206-207-208 for peak 22, and one GAG codon in position 190 for peak 24. The better correlation for peaks 22 and 24 suggests that indeed two different tRNAs may exist for glutamic acid instead of one. The effect of this change in decoding pattern for GAA and GAG was checked elsewhere in the profile but no clear supplementary evidence favouring this proposal could be put forward, because other modifications do not appear in regions where N^* is low. For the same reason, no conclusion could be drawn for the four other codons. Repeated similar analyses for other proteins, with the eventual help of site-directed mutagenesis and insertion of oligonucleotides, could be useful for the clarification of these ambiguities in decoding pattern.

It should be pointed out that the molecular weights of some peaks in the densitometer profile A were accurately determined in the following way. Very short [^{35}S]methionine pulses (Fig. 4, lanes A, B and C) were performed in fully induced cultures of CA31 ColA strain. In regions of the gel where methionine residues were close enough, all intermediates appeared. But, in other regions, only the intermediates immediately downstream from a methionine residue could be seen, while intermediates just upstream could not. Thus, it was possible to determine accurately the real molecular weights of intermediates near methionine residues 185, 262, 295, 387 and 448.

From this comparison between profile B and the profiles A, C and D, the following conclusion can be drawn: in spite of technical difficulties and theoretical problems exposed in the Discussion, the marked correlation that exists between positions of peaks in theoretical and experimental profiles indicates that accumulations of nascent polypeptides are indeed directly related to the tRNA concentrations. This conclusion will be discussed further after analysis of other non-uniform translations.

(d) *Intermediates in synthesis of colicin E1*

The codon usages for the colicin E1 gene (Yamada *et al.*, 1982) and the colicin A gene are rather similar (Morlon *et al.*, 1983); thus similar experimental results were expected for translation of colicin E1 mRNA. Marked intermediates were in fact observed in a pulse-chase experiment (Lazdunski *et al.*, 1984), and the correlation between position of the theoretical and experimental peaks was checked. However, owing to the small number of methionine residues in the protein and to a poor yield of immunoprecipitation in the low M_r range, this correlation is more difficult to establish firmly, and a numbered correspondence between peaks cannot be proposed.

(e) *Intermediates in synthesis of TEM 1-β-lactamase*

The phenomenon of non-uniform translation was investigated *in vivo* for some other proteins in our previous work (Varenne *et al.*, 1982). The theoretical approach described for colicin A was applied to these proteins: TEM 1-β-lactamase encoded by pBR322, the OmpA and LamB proteins, and the elongation factor EF-Tu. Theoretical profiles indicated that marked intermediates could be expected to occur in TEM 1-β-lactamase. As intermediates were not

observed in our previous work, reasons for this discordance were therefore investigated. It turned out that proteolytic degradation of intermediates had occurred under conditions previously used for this protein. New experiments were carried out. As expected from the theoretical approach, accumulation of polypeptides was observed in a 25-second pulse experiment. All these bands disappeared during an 80-second chase, except one that was reinforced and corresponds to mature β-lactamase. The densitometer profile corresponding to the pulse (Fig. 7(a)) was compared to the theoretical profile (Fig. 7(b)). Peak 3 corresponds to mature β-lactamase and peak 1 to pre-β-lactamase, which is quickly processed in β-lactamase only after completion of translation at 37°C (Josefsson & Randall, 1981).

Although the intermediates are more difficult to observe than for colicin A, especially in the low M_r range, the correlation observed between positions of experimental and theoretical peaks confirms that pauses are not created by an imbalance in the pool of the tRNAs, but are related to the physiological differences in concentrations of the tRNAs.

It must be pointed out that if the experimental concentration reported by Ikemura (1981a) for the $tRNA^{Ile}_{minor}$ was used in computations, a reasonable fit could also be observed (Fig. 7(c)) in contrast to results observed with colicin A, for which only the interpolated concentration of this tRNA could be used. This last result suggests that the ColA plasmid might modify the intracellular concentration of this tRNA. If true, this would occur in a permanent way, since no modification of the pattern of pauses could be observed in the experiment described in Figure 4. Further experiments are needed to check this possibility and to evaluate more accurately the concentration of $tRNA^{Ile}_{minor}$ in *E. coli* strains.

(f) *Intermediates in synthesis of pre-OmpA protein*

In synthesis of constitutively highly expressed proteins like EF-Tu (about 7×10^4 copies/cell) and OmpA protein (about 10^5 copies/cell), use of minor tRNAs is avoided. Nevertheless, the other tRNAs do not exist in equal amounts: the ratio of the two extreme tRNA concentrations used in synthesis of these proteins is about 5 instead of about 10 for TEM 1-β-lactamase with the $tRNA^{Ile}_{minor}$ concentration taken as the interpolated value, and 22 with the value from Ikemura (1981a). Faint intermediates would be expected, as shown for the OmpA protein in Figure 9(b). Since a possible secondary structure for the mRNA of OmpA has been proposed (Movva et al., 1980), and since the possibility that hairpins might be involved in discontinuous translation of OmpA was suggested (Varenne et al., 1982), the synthesis of this last protein was analysed and the two hypotheses concerning discontinuous translation were examined.

A typical pulse-chase experiment is shown in Figure 8. Discontinuous translation was again observed with some differences in profile as compared to those for colicin A and TEM 1-β-lactamase.

(1) Pauses were never observed above 23,000 M_r, but this does not mean that translation was uniform in this region: in fact, for an unknown reason, a major part of the growing polypeptide chains was lost in all similar experiments, since a

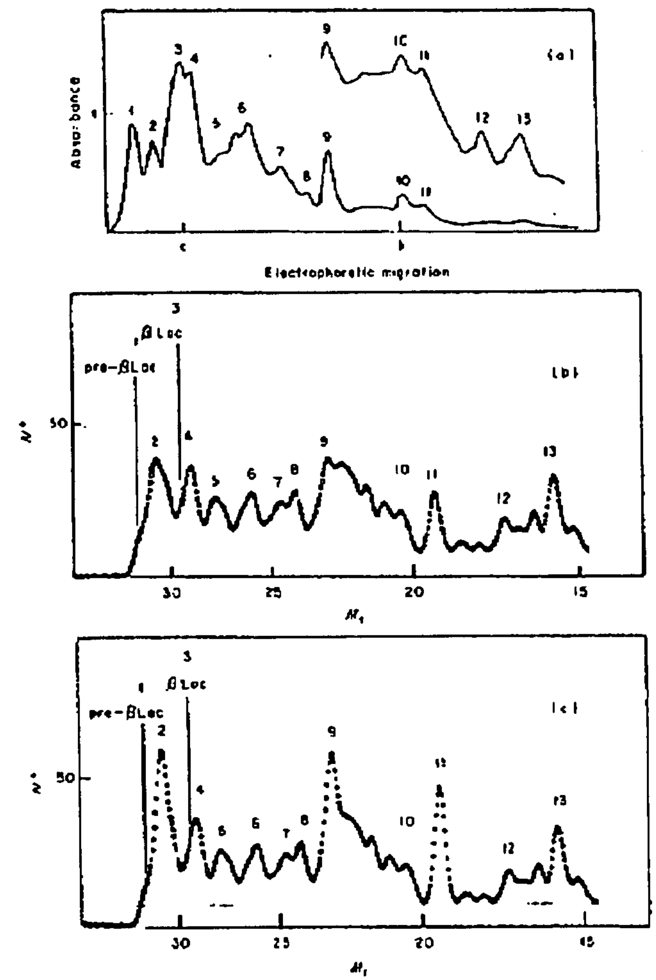


FIG. 7. Comparison of theoretical and experimental profiles for β-lactamase. (Recognition spectrum according to Ikemura, 1981a,b.) (a) Densitometer profile of intermediates. The upper trace corresponds to the scanning of the same gel but exposed 8-fold longer. (b) N^* with FWHM = 11·6 mm. (c) Same as (b) but the experimental concentration of $tRNA^{Ile}_{minor}$ was used. βLac and pre-βLac, β-lactamase and pre-β-lactamase.

uniform translation should produce a continuous background, which was never observed. In any case, interpretation of the upper part of the profile would be difficult for two reasons: first, the abnormal migration of mature OmpA (Nakamura & Mizushima, 1976) and thus probably of high molecular weight intermediates could not be avoided totally; and second, a part of the precursor

form is cotranslationally processed when the size of the nascent polypeptide chain becomes greater than 30,000 M_r (Josefsson & Randall, 1981). Thus, four differently migrating forms are expected for each kind of intermediate.

(2) As expected if pauses are due to tRNA availability, observed accumulations in OmpA synthesis were less marked than in colicins or TEM 1-β-lactamase synthesis, and a much longer exposure of the fluorogram was needed to detect the pauses shown in Figure 10.

(3) Consequently, the background was more important (Fig. 8, lane 2), and must be taken into account for interpretation of the densitometer profile of lane 1 shown in Figure 9(a) and (c). Some bands that do not disappear during the chase, and are also visible in the same chase experiment followed by immunoprecipitation with anti-lipoprotein (lane 4), correspond to abundant proteins of the cell (lane 5) and must be discarded for the analysis of the experimental profile of Figure 9(a) and (c).

Theoretical profiles of N^* are shown in Figure 9(b) and (d). Amino acid residues whose numbers are indicated on the abscissa of Figure 9(d) are those which are bound to tRNA in the ribosome A-site just upstream of the ribosome entry into possible hairpins of the mRNA (Movva et al., 1980). A strong correlation again exists between positions of observed and predicted peaks in the tRNA theory, whereas a full correspondence does not exist in the hairpin theory. Moreover, at least three peaks (2, 3 and 6), corresponding to residues 206, 192 and 164, are found in regions of mRNA where non-optimal codons or non-classified codons (corresponding to His, Asp, Cys or Ser: Ikemura, 1981b) are not encountered.

Two conclusions may be drawn from the above data. First, even for constitutively highly expressed proteins, faint intermediates (which reflect the unequal duration of searches for adapted ternary complexes for the codon in the A-site of the ribosome) can be observed even in parts of mRNA where only "optimal" codons are used. Second, the possible role of mRNA secondary structure in creating a non-uniform rate of translation is probably a minor one, if it exists at all, for the OmpA protein.

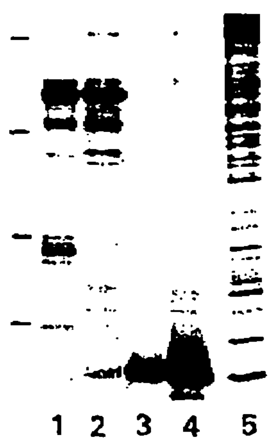


Fig. 8. Intermediates in pre-OmpA synthesis. Cells were pulsed for 30 s with [^{35}S]methionine (lanes 1 and 3) and chased for 120 s (lanes 2, 4 and 5), then solubilized and immunoprecipitated by anti-ompA protein antibody (lanes 1 and 2) or by anti-lipoprotein (lanes 3 and 4). Lane 5, whole cells.

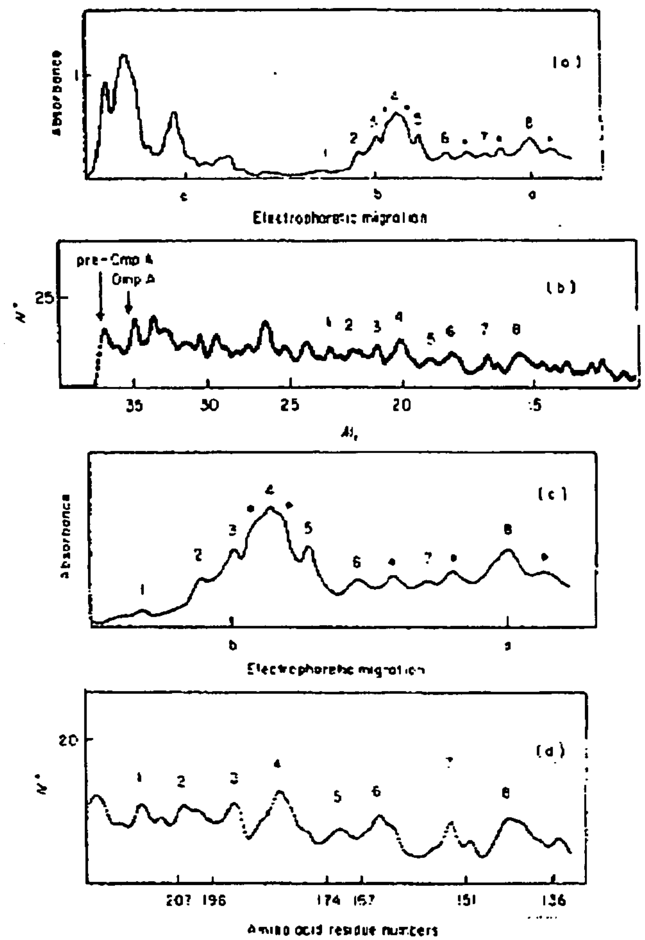


Fig. 9. Comparison of theoretical and experimental profiles for pre-OmpA. (Recognition spectrum according to Ikemura, 1981a,b.) (a) Densitometer profile of intermediates obtained from lane 1 (Fig. 8). The filled circles indicate the bands that do not disappear during the chase, exist in lanes 3, 4 and 5 and do not constitute intermediates of synthesis. (b) N^* with FWHM = 6·4 mm. (c) Partial enlargement of (a). (d) Partial enlargement of (b). The numbers indicated on the abscissa correspond to the amino acids specified by the codon exposed in the ribosomal A-site just upstream of the ribosome entry into possible hairpins in the mRNA (see the text). The numbers of the amino acids corresponding to the tops of the theoretical peaks in the N^* profile are indicated in parentheses. 1 (215), 2 (206), 3 (192), 4 (184), 5 (172), 6 (164), 7 (154), 8 (147).

(g) *Comparison of N^* variations along different mRNAs*

Since the rate of translation is tightly connected to N, it is of interest to know if N (or N^*) presents certain particularities along the polypeptide chain. Observation of the theoretical profiles presented in Figures 5, 7 and 9 does not easily provide this information for two reasons: (1) the N-terminal parts of the proteins are lacking; (2) distortions in the profiles of N^* are introduced by differences in amino acid residues M_r values and by logarithmic migration of polypeptides in the gel.

It was, therefore, more appropriate to plot $N_i = f(x_i)$, where x_i is a linear function of the amino acid residue number and to disperse the results. Three N^* profiles are shown in Figure 10 in order to compare two proteins sharing similar features (colicins A and E1), and to another that is very different from both, though very efficiently expressed (OmpA protein). It must be kept in mind that the choice of the N^* representation instead of that employing N allows a general view of the slowing down regions along the whole mRNA, but minimizes, to an important extent, local variations (see Fig. 6). As mentioned above, the ratio of the two extreme values of N is 10 for both colicins, and 5 for OmpA; 3% of OmpA codons and 17% of colicin A codons lead to N higher than four times the minimal value of N (N minimum = 16·2).

Several remarks can be made about the N^* plots shown in Figure 10.

(1) Regions with a high number of tRNA–codon interactions exist along the whole polypeptide chain and not mostly in the C-terminal part of the molecule, as might be suggested from previous theoretical profiles plotted in a semi-logarithmic representation. Pauses under 13,000 M_r were not detected in whole cells of fully induced CA31 ColA. Two reasons at least could explain this observation: first, intermediate polypeptide chains in this range of M_r were probably very sensitive to proteolytic degradation; secondly, distances between two adjoining polypeptides are approximately threefold smaller below 14,000 M_r than above (Swank & Munkres, 1971). Since the dispersion in the gel remains similar, interference is much more marked between close intermediates.

(2) The variability of N^* during translation of mRNAs is much higher for colicins A and E1 than for the OmpA protein.

(3) $\bar{N}^*$, the average value of N^* calculated over the whole polypeptide chain, is higher for colicins A and E1 than for the OmpA protein. This reflects the difference between $\bar{N}$ values:

$$\bar{N} = \frac{\text{Sum of } N \text{ for the whole protein}}{\text{Number of amino acid residues}},$$

among these proteins.

As a close relationship between high expressivity and low values of $\bar{N}$ was found by Gouy & Gautier (1982), values of $\bar{N}$ were calculated with two decoding patterns (from Ikemura, 1981*a*, and from Grosjean & Fiers, 1982) for the proteins studied in this work (Table 3). To facilitate the discussion, this Table also contains further information: (i) differences in $\bar{N}$ from protein to protein may result, at least partially, from differences in amino acid composition. To take this

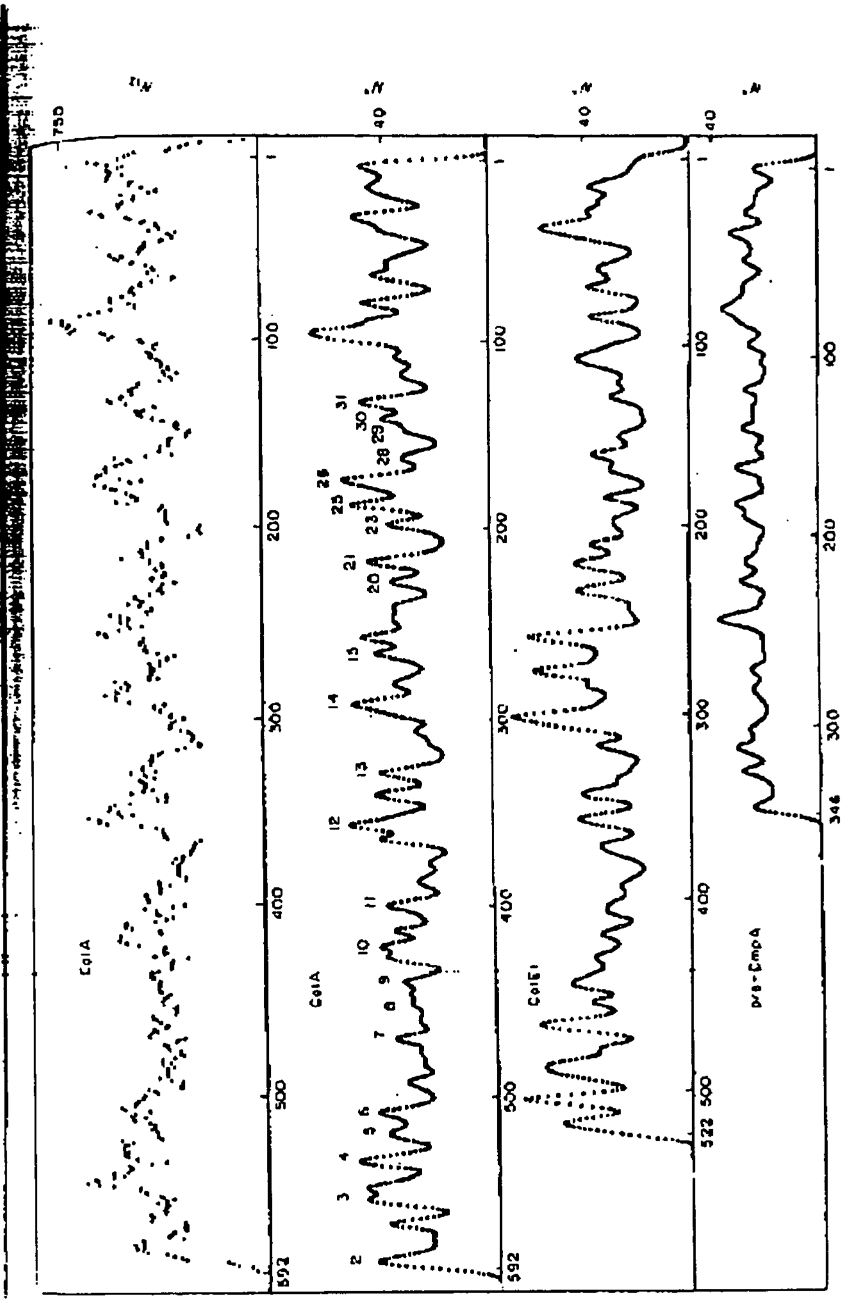


FIG. 10. Comparison between theoretical profiles. From top to bottom: search of ribosome collisions (see the text) for colicin A. Linear plot of N^* with FWHM = 6·6 amino acid residues for colicin A. Linear plot of N^* with FWHM = 5·8 amino acid residues for colicin E1. Linear plot of N^* with FWHM = 5·8 amino acid residues for pre-OmpA protein. Recognition spectrum according to Ikemura (1981*a*,*b*) and interpolated value for $tRNA^{Ile}_{minor}$ were used in the computation.

TABLE 3

Values of the average number ($\bar{N}$) of selections for some mRNAs

Decoding pattern	ColA	ColEI	Bla	OmpA	LamB	PhoE	TufA	LacZ	Average *E. coli*
I	35·6	35·7	42·4	26·9	32·0	33·7	25·1	37·8	29·7
Io	21·4	23·4	25·8	25·3	27·8	28·0	24·3	28·1	24·6
GF	41·8	42·0	47·1	28·5	34·5	37·1	27·6	40·1	32·2
$\frac{I-Io}{Io}$ (%)	34	53	64	6	15	30	3	35	17

In calculation of $\bar{N}$, the interpolated value of the $tRNA^{Ile}_{minor}$ concentration was used. Decoding pattern: I, according to Ikemura (1981a,b); GF, Grosjean & Fiers (1982); Io, Ikemura optimal codon usage.

point into account, the minimal $\bar{N}$ value was calculated with Ikemura's decoding pattern. This optimal $\bar{N}$ value corresponds to an optimal nucleotide sequence using for each amino acid the codon(s) corresponding to the major iso-tRNA; (ii) PhoE, and β-galactosidase are included; (iii) the average tRNA usage from Table 2 allows estimation of $\bar{N}$ for the total *E. coli* cell proteins. These estimates are also included.

(4) If a ribosome pauses too long in a specific region of the mRNA, the next upstream ribosome may also pause because its movement might be impeded by the first. In the model for translation analysed here, regions where such a phenomenon would have a maximum probability of occurring are the 3′ coding extremity of the mRNA if the termination rate is limiting, and in parts of the mRNA where the total average number of selections relative to n adjacent codons (n being the minimal possible distance between two adjacent ribosomes) reaches a maximum value, especially if such a value has not been attained upstream in the mRNA. Thus a diffusion computer program, where the Gaussian distribution was replaced by a unit distribution on n codons, was applied to $\bar{N}$ *versus* residue number. Minimal distances of 12 or 15 codons were used in the kinetic models of protein synthesis presented by Bergmann & Lodish (1979) or Von Heijne *et al.* (1978), respectively. In our program, different values of n between 11 and 17 were used in the theoretical treatment for colicin A. For each codon i, the computer calculates the total average number of trials relative to the next n codons just downstream from codon i (i.e. from codon $i+1$ to codon $i+n$).

We observed that maxima of such a plot dispersed with $n = 12$ (designated as $\bar{N}_{12}$ in Fig. 10) could be predicted from the $\bar{N}^*$ profile shown in Figure 10 (full width at half maximum = 5·8 residues), and particularly that the most important one was located, as in the $\bar{N}^*$ profile, around residue number 95. The other values of n led to similar conclusions about positions of the maxima. In fact, such "secondary pauses" were not detected in scanning intermediates, even in regions corresponding to maxima of the $\bar{N}_{12}$ plot. It is of interest that the 12 first codons lead to a high value of $\bar{N}_{12}$, which may interfere with the initiation rate of translation for colicin A mRNA.

4. Discussion

Two important conclusions can be derived from our results: (1) the rate-limiting step in the elongation cycle of polypeptide chains is the search for the ternary complex (aminoacyl-tRNA bound to EF-Tu and GTP) specific to the codon at the A-site of the ribosome. The next two steps in the cycle, that is transpeptidation and translocation, account for a much shorter time than the delay before the successful collision with the specific iso-tRNA; (2) inasmuch as tRNAs are not in equimolar concentrations in the cell cytoplasm, including those (the most abundant) that are preferentially used for highly expressed proteins of *E. coli*, the elongation of polypeptide chains must occur at a variable rate for all *E. coli* proteins.

We had to overcome a number of technical difficulties in order to interpret our experimental data. The first was related to the heterogeneous distribution of methionine residues, which might result in heterogeneous labelling along the polypeptide chain. The second dealt with the possible loss of material before immunoprecipitation. The third problem came from a yield of immunoprecipitation that might not be constant for all intermediates. The most serious difficulty was a possible proteolytic degradation of intermediates that could not always be suppressed. This difficulty appeared especially in TEM 1-β-lactamase experiments, and seemed to be highly variable from protein to protein. By shortening each step when possible, we could alleviate this problem. However, future use of protease mutants such *lon* (Grossman *et al.*, 1983) and/or the use of a protease inhibitor might further alleviate this problem.

All the difficulties evoked above were circumvented for colicin A. However, theoretical problems remain.

(1) The logarithmic migration for intermediates can suffer considerable local deviations (see colicin A intermediates, Fig. 5).

(2) A constant dispersion was used for computation, although the dispersion along the gels is clearly not constant.

(3) Experimental tRNA concentrations are known with a non-neglectable standard deviation, especially for minor tRNAs.

(4) The decoding spectrum is not known with certainty for a certain number of codons, and the determination of "apparent concentrations" when a codon is partially recognized *in vitro* by a second tRNA is questionable because of lack of adequate data obtained *in vivo*.

(5) The accuracy in concentration for those cases obtained by interpolation might be lower than that for those which have been determined experimentally for two reasons: (i) codon usage, although it was established from 62 genes, can only be approximate; and (ii) an exactly linear relation is not likely to exist between frequency of tRNA usage and amount of tRNA. The following example gives an idea of the uncertainties introduced by interpolation. According to Ikemura (1981a), $tRNA^{Trp}$ and $tRNA^{Met}_{m}$ have the same concentration, but Table 1 indicates that their usage is quite different. Thus, interpolated values of concentrations (1·13% and 2·33%) would lead to numbers of discriminations (88 and 43) very different from that (60) deduced from the experimental concentration.

(6) Differences between tRNA frequencies and operational frequencies of ternary complexes might exist.

(7) $\bar{N}_i$ was calculated as $1/C_i$ but the probability of transpeptidation after each collision between a codon and a cognate-tRNA is probably not 1 for each tRNA, and might differ from one tRNA to another; this would introduce increases of $\bar{N}_i$ and distortions in the profile if the percentage increase is not the same for all codons.

(8) The mean duration for the addition of an amino acid corresponding to a given codon may be obtained from $\bar{N}$ by the simple relation $t = (\theta_0 + \theta_1)\bar{N} + t_2 + t_3$ only if θ_0, t_2 and t_3 are the same for all codon species (θ_1 is independent of the codon). In fact, this is probably approximate. For example, any attempt to explain codon usage should involve both tRNA concentrations and a consideration of the energetics of codon–anticodon pairing: one tRNA can often translate two codons and a bias in the codon usage has been observed (Grosjean & Fiers, 1982; Grantham *et al.*, 1981; Ikemura, 1981*a*,*b*). This bias might play a role in fidelity and/or in the rate of translation of codons. There might thus exist an effect of codon choice on translation rate by modification of θ_0, t_2 and t_3 that we could not take into account for lack of information.

Experimentally, the role of codon choice could not be demonstrated in colicin A in a quantitative way. For example, non-optimal codons (which, as a general rule, are rarely used in constitutively highly expressed genes) with a low energy of interaction are found between peaks 13 and 14 in positions 311 and 314 for AAT (Asn) and 313 for ATT (Ile). These codons do not seem to induce an additional slowing down of ribosomes in this region (see Fig. 5), but it is not possible to draw general conclusions about these codons from one example. Elsewhere, we cannot rule out the hypothesis that non-optimal codon usage leading to high energies of interaction might have resulted in additional slowing down of ribosomes.

These theoretical and experimental difficulties probably explain why areas of experimental and theoretical profiles do not match exactly, even for colicin A.

TABLE 4

Rates of synthesis of various proteins in E. coli

Protein	$\bar{N}$	Rate A (amino acids/s)	Protein	$\bar{N}$	Rate B (amino acids/s)
EF-Tu	25·1	15·3–20·4	EF-Ts	21·6	5·6
OmpC/F protein	27·0 and 28·1	12·1–15·2	S_1 (ribosomal)	25·0	4·3
			EF-Tu	25·1	4·8
OmpA protein	29·9	10·9–13·2	EF-G	≃27	5·0
LamB protein	32·0	14·8–17·8	*lac* repressor	35·8	3·4
ampC β-lactamase	39·9	<9·5	Tem I-β-lactamase	42·4	3·4
Tem I-β-lactamase	42·4	7·4–9·8			

The rates of assembly of amino acids were assayed: rate A, at 37°C by Josefsson (1982); rate B, at 24·5°C by Pedersen (1983). $\bar{N}$ was computed as in the first line of Table 3. References for nucleotide sequences of proteins presented above are cited in the text. The value for EF-G is approximate, since the nucleotide sequence is incomplete.

However, a very significant overall correspondence was observed between positions of peaks, which indicates that experimentally observed variations in elongation rates have their main origin in tRNA availabilities. Moreover, the importance of the gap between maxima and minima in the experimental profiles means that the value of B in the equation $t = A\bar{N} + B$ is low compared to $A\bar{N}$. This provides a direct demonstration that transpeptidation and translocation steps occupy a short or neglectable time as compared to the aminoacyl-tRNA selection step. This conclusion is strongly supported by the experimental determination of rates of translation for different mRNAs performed by Josefsson (1982) and Pedersen (1983). The marked decrease of the rates of translation observed when $\bar{N}$ is high (*ampC* β-lactamase, TEM I-β-lactamase, *lac* repressor: see Table 4) confirms the prevalence of the discrimination step in the elongation cycle.

However, we cannot exclude the possibility of a modulation of a rate of translation by the energetics of codon–anticodon pairing. Furthermore, additional factors like mRNA secondary structure might also contribute to discontinuous translation for proteins such as MS2 coat protein (Min Jou *et al.*, 1972; Chaney & Morris, 1978) that we have not studied.

Experiments on discontinuous translation may lead to valuable new or supplementary information concerning translation *in vivo*, concerning, for example: the recognition pattern of certain tRNAs; tRNA concentrations; the possible effect of codon–anticodon energies of interaction; the possible influence of codon context. Site-directed mutagenesis and insertion of oligonucleotides should allow a more direct approach to these problems.

It is necessary to enlarge the conclusion that the elongation of polypeptide chains occurs at variable rate for *E. coli* proteins. This means that for one mRNA species and for one given codon, the duration of addition for the corresponding amino acid residue fluctuates around an average value. From this codon to the next one upstream, this average value varies in a ratio rather similar to the inverse ratio of the tRNA concentrations corresponding to these codons. Thus, the average rate of assembly of amino acids for a given protein is approximately proportional to the inverse of the average value $\bar{N}$ of the selection numbers for the whole protein, but individual rates for individual mRNAs vary (i.e. each individual mRNA is not translated at the same rate for a given protein). This implies that comparison between rates of translation for two proteins must be performed for the same parameter: average rate, minimal detectable rate, maximum detectable rate.

At least four experimental observations argue for dispersion of translation rates in individual mRNAs.

(1) In the experiment shown in Figure 4 for determination of the exact location of some intermediates, radiolabelling of intermediates after a 30-second chase is shown in lane D. If all individual translational rates were identical, the nascent chains that were upstream from methionine residue 202, for example, at the beginning of the chase and were not labelled (appearing as a blank area in lane C) should be longer 30 seconds later, and migrate in the gel like polypeptides of 360 to 400 residues. In fact, radiolabelling was present in this area, indicating

differences in individual rates of amino acid assembly (as judged by radiolabelling intensities, delay in chase could not alone account for the observed result).

(2) As mentioned above, a [^{35}S]methionine pulse for 20 seconds provided a labelling of intermediates in synthesis of colicin A rather similar to that obtained with a mixture of ^{14}C-labelled amino acids, in spite of an irregular distribution of methionine residues along the polypeptide chain. The duration of pulse labelling could not alone account for this fact if the elongation rate was the same for all mRNAs.

(3) In colicin E1 experiments, the translation rate deduced from the appearance of [^{35}S]methionine in mature colicin E1 (calculated on 152 residues, since the C-terminal methionine residue is in position 370) is higher than the translation rate deduced from the disappearance of pauses (not shown). The same conclusion can be drawn from the colicin A experiments.

(4) For β-galactosidase ($\bar{N} = 38$ as for colicin A), a variable rate of translation, ranging from 8 to 15 amino acids per second has been reported (Talkad *et al.*, 1976).

This dispersion of individual rates of translation explains why the minimum detectable translation rate deduced for colicin A (592 amino acids translated in about 70 to 80 s leading to an approximate value of 8 residues/s: Varenne *et al.*, 1982) must be compared with the minimal value for β-galactosidase, and not with the maximal value deduced from appearance of enzymatic activity after induction. It is significant that the minimum values for colicin A and β-galactosidase are similar, as both proteins have the same $\bar{N}$ value, and similar values for the P_2 index from Gouy & Gautier (1982), which characterizes the choice between codon–anticodon pairing energies (0·49 for colicin A, 0·45 for β-galactosidase).

In agreement with previous studies (Gouy & Gautier, 1982), the two constitutively highly expressed proteins in Table 3 have an $\bar{N}$ value close to the optimal value, since deviations between $\bar{N}$ and $\bar{N}$ optimal are 3% for EF-Tu and 6% for OmpA protein, in contrast with other proteins in this Table (35% for β-galactosidase, 54% for colicins A and E1, 64% for TEM 1-β-lactamase). This confirms once more that constitutively highly expressed proteins are encoded by genes highly adapted to the tRNA content of the cell for fast translation. It is of interest to observe that the degeneracy of the genetic code introduces an important potential variability in possible N values. If the less abundant iso-tRNAs corresponding to each amino acid were exclusively used in synthesis of colicin A or OmpA protein, for example, the average number of selections should attain values of 65 and 68, respectively; i.e. about 2·7-fold the minimal possible value (about 3·3-fold the minimal possible value if the recognition pattern according to Grosjean & Fiers (1982), and the experimental value of $tRNA^{Ile}_{minor}$ are used in the calculations). Among known nucleotide sequences of proteins synthesized in *E. coli*, the highest values are never approached, in contrast to the lowest values. The highest value among known nucleotide sequences is 61·2 for the immunity protein for ColE1 (maximum possible value, 77·3: Lloubes *et al.*, 1984).

One can address the question of the physiological significance of discontinuous

translation and ask whether it is a mere reflection of different tRNA concentrations in the cytoplasm or whether it reflects any particular regulatory strategy of the cell. It is now well-established that highly expressed mRNAs generally use abundant tRNAs and "optimal" pairing energies, and that weakly expressed mRNAs often display an opposite choice (Grantham *et al.*, 1981: Grosjean & Fiers, 1982), leading to a slower translation. It is quite possible that local variations in elongation rate along mRNAs have no physiological finality, and that the essential objective of more or less marked general slowing down is to introduce a constitutive modulation in average rates of translation for the mRNAs, according to cellular needs (Gouy & Gautier, 1982). Besides, since transcription and translation are coupled, it is possible that transcription also occurs at different average rates for different genes and that the use of rare iso-tRNAs merely adapts translation to slow transcription. The fact that the same intermediates were also observed when transcription was blocked (Varenne *et al.*, 1982) does not exclude the possibility of the existence of such an adjustment between rates of transcription and translation.

However, local variation in translation rates, at least in specific cases, might have a physiological significance, for example by favouring sequential polypeptide chain folding. Variable rates of polypeptide elongation might allow short and medium-range interactions to take place before long-range interactions in the polypeptide chains, thus favouring domain formation. With regard to this point, it should be recalled that colicins have well-structured domains (De Graaf *et al.*, 1978; Ohno-Iwashita & Imahori, 1980). Furthermore, tight coupling between transcription and translation might be necessary or advantageous to ensure efficient synthesis. Discontinuous transcription, which has been reported in specific cases (Darlix & Fromageot, 1972; Yanofsky, 1981; Kingston & Chamberlin, 1981), might be a more general phenomenon, and variations in translation rate might be part of a tight coupling mechanism between transcription and translation.

These problems bear on molecular biology but also obviously bear on biotechnology. It is desirable to use a microorganism having a tRNA pool as much adapted as possible to the message being translated, or if the gene is an artificial one, to use only optimal codons from the host organism, if the only significance of the phenomenon described is to globally attune syntheses; this should allow highly expressed proteins to be produced as rapidly as possible, and/or as faithfully as possible and/or as economically as possible. However, efficiency of production of a protein is not always related to an optimal codon usage. With regard to this point, the case of β-galactosidase is particularly illustrative. Although $\bar{N}$ is rather high ($\bar{N} = 38$) for this enzyme, which probably leads to the premature termination *in vivo* and *in vitro* observed by Manley (1978), a high level of production is obtained (about 3% of total cell proteins). A similar situation exists for colicin A, for which a very high level of synthesis coexists with premature termination *in vivo* and *in vitro* at pause sites (unpublished results). This suggests that a high expression is not necessarily related to a maximal rate of translation, and that the choice of codons in artificial genes, for example, need not always be restricted to the optimal ones.

Colicin A and E1 are highly expressed proteins after induction (even natural induction), but have a codon usage different from that of highly expressed proteins of *E. coli*. One hypothesis suggested that it was the plasmid-coded nature of colicin genes that was responsible for their codon usage not being adapted to the tRNA population of *E. coli* (Varenne *et al.*, 1982). However, from nucleotide sequences published recently it can be deduced that chromosomally encoded inducible proteins that are highly expressed, like PhoE protein (Overbeeke *et al.*, 1983) and β-galactosidase (Kalnins *et al.*, 1983), also have a codon usage quite different from that of proteins like ribosomal proteins (see Table 3 for $\bar{N}$). It appears, therefore, that among highly expressed proteins, perhaps constitutive ones should be distinguished from inducible ones with regard to codon usage. Then RecA protein should be considered as a constitutive protein (1000 copies/cell: Karu & Belk, 1982), although its synthesis can be induced to higher levels.

Little information about discontinuous elongation in eukaryotic cells has been reported. However, in at least two cases, for fibroin (Lizardi *et al.*, 1979; Chavancy & Garel, 1981) and for globin (Protzel & Morris, 1974; Chaney & Morris, 1978), non-uniform peptide elongation was clearly demonstrated *in vitro* and *in vivo*. The specific case of fibroin is especially interesting because of the peculiar amino acid composition of this protein in which Ala, Gly and Ser residues account for 87% of the amino acid residues. The tRNA population in the posterior silk gland is well-adapted to this amino acid composition and leads to a fast translation of the message. However, clear pauses in translation were observed that probably correspond to the existence of stretches of rare codons intercalated between repetitive domains rich in glycine, alanine and serine (Chavancy & Garel, 1981). Moreover, these results show that there is no contradiction between discontinuous and efficient translation.

In conclusion, we presume that also in other prokaryotic and eukaryotic organisms the stochastic search of the ternary complex specific to the codon at the A-site of the ribosome leads to a non-uniform translation. Further experiments will be carried out in our laboratory to confirm this point.

During the course of this work, discussions with Drs R. Grantham, M. Gouy, R. Buckingham, M. Springer, J. Ninio and D. Cavard were very helpful. We specially thank Dr R. Buckingham for careful reading of the manuscript. We are grateful to M. Knibiehler for excellent technical assistance and to A. Journet for helping in computation. We thank Drs B. Lugtenberg, M. Schwartz, G. Cesarini and J. Rosenbuch for generous gifts of specific antisera. This research was supported by grants from the ATP "Biologie Moléculaire du Gène" and "Microbiologie" and from the Fondation pour la Recherche Médicale.

REFERENCES

Aiba, H., Fujimoto, S. & Ozaki, N. (1982). *Nucl. Acids Res.* 10, 1345–1361.
Ames, B. N. & Hartman, P. E. (1963). *Cold Spring Harbor Symp. Quant. Biol.* 28, 349–356.
Anderson, W. F. (1969). *Proc. Nat. Acad. Sci., U.S.A.* 62, 566–573.
Bergmann, J. E. & Lodish, H. F. (1979). *J. Biol. Chem.* 254, 11927–11937.
Chaney, W. G. & Morris, A. J. (1978). *Arch. Biochem. Biophys.* 194, 283–291.
Chavancy, G. & Garel, J. P. (1981) *Biochimie*, 63, 187–195.

Clément, J. M. & Hofnung, M. (1981). *Cell*, 27, 507–514.
Cossart, P. & Gicquel-Sanzey, B. (1982). *Nucl. Acids Res.* 10, 1363–1378.
Cossart, P., Katinka, M. & Yaniv, M. (1981). *Nucl. Acids Res.* 9, 339–347.
Darlix, J. L. & Fromageot, P. (1972). *Biochimie*, 54, 47–53.
Deeley, M. C. & Yanofsky, C. (1981). *J. Bacteriol.* 147, 787–790.
De Graaf, F. K., Stukart, M. J., Boogerd, F. C. & Metselaar, K. (1978). *Biochemistry*, 17, 1137–1142.
Gouy, M. (1981). Thèse de Troisième Cycle, Université de Lyon I.
Gouy, M. & Gautier, C. (1982). *Nucl. Acids Res.* 10, 7055–7074.
Gouy, M. & Grantham, R. (1980). *FEBS Letters*, 115, 151–155.
Grantham, R., Gautier, C. & Gouy, M. (1980). *Nucl. Acids Res.* 8, 1893–1912.
Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981). *Nucl. Acids Res.* 9, r43–r74.
Grosjean, H. & Fiers, W. (1982). *Gene*, 18, 199–209.
Grossman, A. D., Burgess, R. G., Walter, W. & Gross, C. A. (1983). *Cell*, 32, 151–159.
Grundstrom, T. & Jaurin, B. (1982). *Proc. Nat. Acad. Sci., U.S.A.* 79, 1111–1115.
Hall, C. V., Van Cleemput, M., Muench, K. H. & Yanofsky, C. (1982). *J. Biol. Chem.* 257, 6132–6136.
Ikemura, T. (1981*a*). *J. Mol. Biol.* 146, 1–21.
Ikemura, T. (1981*b*). *J. Mol. Biol.* 151, 389–409.
Ikemura, T. (1982). *J. Mol. Biol.* 158, 573–597.
Josefsson, L. G. (1982). Ph.D. thesis, University of Uppsala.
Josefsson, L. G. & Randall, L. L. (1981). *Cell*, 25, 151–157.
Joyce, C. M., Kelley, W. S. & Grindley, N. D. F. (1982). *J. Biol. Chem.* 257, 1958–1964.
Kalnins, A., Otto, K., Rüther, V. & Müller-Hill, B. (1983). *EMBO J.* 2, 593–597.
Karu, E. A. & Belk, E. D. (1982). *Mol. Gen. Genet.* 185, 275–282.
Kingston, R. E. & Chamberlin, M. J. (1981). *Cell*, 27, 523–531.
Laskey, R. A. & Mills, A. D. (1975). *Eur. J. Biochem.* 56, 335–341.
Lazdunski, C., Morlon, R., Lloubès, R., Varenne, S., Knibiehler, M., Chartier, M., Bernadac, A. & Cavard, D. (1984). In *Bacterial Protein Toxins* (Alouf, J., ed.), pp. 107–113, Academic Press, London.
Lin, S. & Zabin, I. (1972). *J. Biol. Chem.* 247, 2205–2211.
Lizardi, P. M., Mahdavi, V., Shields, D. & Candelas, G. (1979). *Proc. Nat. Acad. Sci., U.S.A.* 76, 6211–6215.
Lloubès, R., Chartier, M., Journet, A., Varenne, S. & Lazdunski, C. (1984). *Eur. J. Biochem.* 144, 73–78.
Manley, J. L. (1978). *J. Mol. Biol.* 125, 407–432.
Min Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. (1972). *Nature (London)*, 237, 82–88.
Morlon, J., Lloubès, R., Varenne, S., Chartier, M. & Lazdunski, C. (1983). *J. Mol. Biol.* 170, 271–285.
Movva, N. R., Nakamura, K. & Inouye, M. (1980). *J. Mol. Biol.* 143, 317–328.
Nakamura, K. & Mizushima, S. (1976). *J. Biochem.* 80, 1412–1422.
Ohno-Iwashita, Y. & Imahori, K. (1980). *Biochemistry*, 19, 652–659.
Ovchinnikov, Y. A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, B. O., Salomatina, I. S., Shuvaeva, T. M., Lipkin, V. M. & Sverdlov, E. D. (1982). *Nucl. Acids Res.* 10, 4035–4044.
Overbeeke, N., Bergmans, H., Van Mansfeld, F. & Lugtenberg, B. (1983). *J. Mol. Biol.* 163, 513–532.
Pedersen, S. (1983). *Alfred Benzon Symp.* 19, 101–111.
Protzel, A. & Morris, A. J. (1974). *J. Biol. Chem.* 249, 4594–4600.
Putney, S. D., Royal, N. J., De Vegvar, H. N., Herlihy, W. C., Biemann, K. & Schimmel, P. (1981). *Science*, 213, 1497–1500.
Smith, D. R. & Calvo, J. M. (1980). *Nucl. Acids Res.* 8, 2255–2274.
Swank, R. T. & Munkres, K. D. (1971). *Anal. Biochem.* 39, 462–477.

Talkad, V., Schneider, E. & Kennell, D. (1976). J. Mol. Biol. 104, 299–303.
Varenne, S., Cavard, D. & Lazdunski, C. (1981). Eur. J. Biochem. 116, 615–620.
Varenne, S., Knibiehler, M., Cavard, D., Morlon, J. & Lazdunski, C. (1982). J. Mol. Biol. 159, 57–70.
Von Heijne, G., Nilson, L. & Blomberg, C. (1978). Eur. J. Biochem. 92, 397–402.
Von Wilcken-Bergmann, B. & Muller-Hill, B. (1982). Proc. Nat. Acad. Sci., U.S.A. 79, 2427–2431.
Yamada, M., Ebina, Y., Miyata, T., Nakazawa, T. & Nakazawa, A. (1982). Proc. Nat. Acad. Sci., U.S.A. 79, 2827–2831.
Yanofsky, C. (1981). Nature (London), 289, 751–757.

Edited by P. Chambon